SDC-0055-057-LF-5028A
May 2, 1996

STANDARD REFERENCE TABLES (SRT) SUMMARY REPORT

WORKING PAPER

CONTRACT NO. 68-W1-0055
DELIVERY ORDER NO. 057

Prepared for:

United States Environmental Protection Agency
Office of Information Resources Management
Program Systems Division
401 M Street, SW.
Washington, DC 20460

Delivery Order Project Officer:

Larry Fitzwater

Prepared by:

EPA Systems Development Center
(A Contractor Operated Facility)
Science Applications International Corporation
200 North Glebe Road, Suite 300
Arlington, VA 22203

CONTENTS

EXHIBIT

EXECUTIVE SUMMARY

This summary report is to announce the availability of an initial set of standard reference domains for use in all information systems within the Environmental Protection Agency (EPA). Standard reference domains provide a single source of standard data elements and data values that are frequently used in environmental information systems. They can be used as access points to environmental data and used to link data between information systems.

The Information Resource Management (IRM) Executive Steering Committee (ESC) of EPA is sponsoring this effort to design and build Standard Reference Domains, implemented as Standard Reference Tables (SRTs), for use in information systems throughout the Agency. Use of SRTs make environmental data easier to access across systems, easier to use, and reduces the cost of data acquisition and system development. SRTs contain frequently used information such as names, codes, and abbreviations for states, counties, countries, and Indian Tribes. SRTs also provide standardized programmatic and scientific data elements and data values, such as hydrologic unit codes and chemical identifiers. Previously, this information was often obtained redundantly by each system development effort and implemented in dissimilar ways in different systems. This made data acquisition more costly and made it very difficult to pull related data together from the dissimilar systems.

The Office of Information Resources Management (OIRM), Enterprise Information Division (EIMD), led the effort. The design, implementation, and testing was supervised by the OIRM Enterprise Systems Division (ESD) and performed as a joint effort in the Information and Data Management Service Center (IDMSC) North at the EPA Systems Development Center (SDC) and by the IDMSC South at Research Triangle Park (RTP). The IDMSC was established by OIRM to provide assistance to all EPA offices in standardizing and managing environmental data. The Office of Water took the lead role in defining the requirements, needs, and interests of system developers and users and in performing the initial testing. Information system managers throughout the Agency were consulted in the initial phase of the effort.

The initial set of SRTs are available to all EPA information system managers and system developers and can be accessed through direct or network connections to the mainframe. The content of the SRTs should be useful for many EPA system development and maintenance efforts. The initial utilization is voluntary and will be used to test and validate the design and access features implemented for the SRTs.

Federal Information Processing Standard Publications (FIPS PUB) are used as the source of SRTs data where appropriate. Other standard data sources are also used where they are the most authoritative (e.g., Bureau of Census, which is the source of tribal land data for FIPS PUB 55).

Future work on this project will develop notification and other procedures for change management. Future work will also extend the tables to a broader coverage to build a common core of key data elements used in EPA information systems.

The Points of Contact (POC) for obtaining more information about using the SRTs, whether for direct access or for implementation in another operational environment, are Larry Fitzwater or Jerry Widdowson, both at EIMD mail code (3405R). Both of the POC can also be contacted through EPA E-mail.

1.0 INTRODUCTION

This summary report is to announce the availability of an initial set of standard reference domains for use in all information systems within the Environmental Protection Agency (EPA). Standard reference domains provide a single source of standard data elements and data values that are frequently used in environmental information systems. They can be used as access points to environmental data and to link data between information systems.
The Information Resource Management (IRM) Executive Steering Committee (ESC) of EPA is sponsoring this effort to design and build Standard Reference Domains, implemented as Standard Reference Tables (SRTs), for use in information systems throughout the Agency. Use of SRTs make environmental data easier to access across systems, easier to use, and reduces the cost of data acquisition and system development. SRTs contain frequently used information such as names, codes, and abbreviations for states, counties, countries, and Indian Tribes. SRTs also provide standardized programmatic and scientific data elements and data values, such as hydrologic unit codes and chemical identifiers. Previously, this information was often obtained redundantly by each system development effort and implemented in dissimilar ways in different systems. This made data acquisition more costly and made it very difficult to pull related data together from the dissimilar systems.

The Office of Information Resources Management (OIRM), Enterprise Information Division (EIMD), led the effort. The design, implementation, and testing was supervised by the OIRM Enterprise Systems Division (ESD) and performed as a joint effort in the Information and Data Management Service Center (IDMSC) North at the EPA Systems Development Center (SDC) and by the IDMSC South at Research Triangle Park (RTP). The IDMSC was established by OIRM to provide assistance to all EPA offices in standardizing and managing environmental data. The Office of Water took the lead role in defining the requirements, needs, and interests of system developers and users and in performing the initial testing. Information system managers throughout the Agency were consulted in the initial phase of the effort.

2.0 BENEFITS

This central, standard source of reference data provides the following benefits to the Agency and EPA's data users:

- Data integrity and consistency are improved. Accurate, current reference data are always available for use by EPA programs and program systems.

- Data sharing for system integration will have improved feasibility. All Agency program systems can use the same data values to represent data elements commonly used by EPA programs.

- Procurement and maintenance of reference data is more convenient. SRTs remove the necessity for each program system to procure and maintain its own reference data.

- Cost to the Agency can be reduced. Only one set of SRTs data sources need to be acquired.

- Need for separate, redundant data tables is reduced. Program systems can access SRTs directly for data validation and display. SRTs can also be used by duplicating the tables on other servers or by downloading flat files to implement SRTs in other database environments.

- The concurrency effect of data issues is minimal. SRTs use an Internal Serial Number (ISN) as the primary key in all reference data tables. The ISN will always be assigned to the same geopolitical or other unit and will track back to all data values assigned to that unit (e.g., as changes are made to a FIPS code or name). Valid information is always available without the need to change system data where an ISN is used as the primary key.

- Impact on users will be minimal. Although application developers are encouraged to use the ISN as the code for standard data, traditional codes (e.g., FIPS codes) are included in the SRTs for those who choose to use them instead.

- Access is provided to current and prior data. The SRTs approach to versioning enables users to use current reference data values in addition to providing the capability to access superseded reference data values.

- A consistent implementation method is used throughout the SRTs. The SRTs implementation approach can be used consistently for all types of SRTs (e.g., test methods and identification of regulated substances in addition to geopolitical standards).

- The SRTs' data structure enables system developers to access data through "views" that can reconstruct the reference table for any point in time, eliminating the need for versioning.

- An audit trail is maintained for each SRTs. Each SRT includes an audit table that tracks data changes, when they occurred, and the authority for changing the data.

- Comprehensive metadata that documents data sources and operational history is provided. Metadata are included at the record level for all data changes and at an operational level for bibliographic information and implementation history for all data loads and updates. Users will be notified of changes to any of the SRTs through automated announcements and through direct notification to all known users.

- SRTs are simple to maintain. The SRTs' implementation eliminate the need for versioning and to identify releases, since any version can be obtained through views. OIRM maintains the tables and provides them to the users through direct access or downloaded files that can be used to refresh tables according to users' needs.

3.0 DATA CONTENT

The SRTs currently include five domains of geopolitical data, including:

- Countries of the world.

- Primary administrative subdivisions of countries. Initially, only the states and provinces of the United States, Canada, and Mexico are included.

- Counties. Initially, only the primary subdivisions of states of the United States (U.S.), its possessions, and freely associated territories of the U.S. are included.

- Tribal lands (i.e., American Indian and Native Alaskan territories).

- EPA regions, cross-referenced to the states that comprise the regions.

In addition, the SRTs contain metadata that provide operational history and bibliographic information about the data.

A detailed logical data model is attached to this document as Appendix A. An overview of the data content is provided as text in the following subsections.

3.1 Country

The country table includes all of the countries of the world. Every country is represented by a 2-character alphabetic code and the name of the country.

3.2 Primary Administrative Subdivision

The primary administrative subdivision table for the initial SRT implementation includes states and state equivalents of the U.S., U.S. possessions and associated territories, provinces of Canada, and states of Mexico. Subdivisions of countries outside of North America are not included in this release.

This table includes FIPS numeric codes and names that represent each of the states and provinces. The table contains alphabetic codes used by the U.S. Postal Service (USPS) as additional information about U.S. states and its related territories. No alphabetic codes have been identified for Canadian provinces and Mexican states.

Foreign keys in this table relate the states of the U.S. and state equivalents to the 10 EPA regions.

3.3 County Table

The county table contains codes and names for counties and county equivalents within the states of the U.S. and state equivalents, including U.S. possessions and freely associated territories. No other countries are represented in the county table. The table includes the FIPS numeric county code and the county name.

3.4 Region Table

The EPA region table contains codes and names for the 10 EPA regions.

3.5 Tribal Lands Tables

The tribal lands data are contained in two relational tables. One table contains the name of the tribal land and its code, as assigned by the U.S. Bureau of the Census, Code 90. The second table provides a cross-reference of that tribal land to the state(s) and counties where the tribal land is located.

3.6 Metadata Tables

Two metadata tables contain data about SRT data: the bibliography table and the operating history table. Each document (e.g., FIPS Publication, ANSI standard, ISO standard, etc.) that is a source of SRTs data is described in the bibliography table with publication number, name, date of publication, author or responsible institution, and primary data sources. The operating history table links each data field in each table to the published source for that data.

4.0 DATA STRUCTURE

Each domain consists of a base table and an audit table. Both tables are dynamic. All records in the base table are assigned a unique, non-intelligent number (i.e., an ISN), which is used to link the base table with the audit table and to link the base tables to related tables (e.g., to link countries to states and provinces). The ISN ensures data integrity across all SRTs, and, where used by program systems as the primary key for data reference, it ensures that all data changes are transparent to the user. Use of the ISN is mandated for SRTs. System developers are encouraged but not required to use the ISN as keys to data values.

The base table always contains the most current data values with a status flag to indicate whether the record was added, changed, or deleted, and the date of the last transaction, using a time stamp. The audit table contains all previous, superseded data values and a time stamp for when the values changed, keyed on ISN and time stamp. The audit table enables tracking of changes to standard data, including code changes, name changes, and deleted values. Views can recreate tables that were valid for a defined time frame, by combining tables and selecting records based on transaction date and record-level data status.

In addition to the ISN, which is used as the primary key for each SRT, these tables contain natural keys. The natural keys are redundant for the SRTs. They are included to satisfy the needs of any program system to use natural keys and to perform ad hoc queries of the reference tables.

5.0 DATA SOURCES

Data sources for each of the SRTs are identified in the Publication and Date column in Exhibit 1. The publication date listed for FIPS PUB 6-4 is the date of the most recent change notice (i.e., change number 3).

6.0 LOCATION OF SRTs

The SRTs are implemented as DB2 and as ORACLE tables on the EPA's IBM mainframe at the National Computer Center. Program systems can use the SRTs through direct access.

System developers who prefer to download SRT data for implementation (e.g., program systems using other database management systems) are encouraged to create and download flat files of SRT data as needed. System developers may choose to create views that reproduce an SRTs at a given point in time.

The entire set of SRTs should be downloaded when program system managers want to refresh their data. Both base tables and audit tables should be downloaded to enable access to superseded data in addition to current data.

Table Name	Publi-cation and Date	Title of Publication	Comments
Country	FIPS PUB 10-4, 1995 April	Countries, Dependencies, Areas of Special Sovereignty, and their Principal Administrative Divisions.	All country codes and names except those for U.S. possessions and freely associated territories.
Primary Subdivision	FIPS PUB 10-4 1995 April	Countries, Dependencies, Areas of Special Sovereignty, and their Principal Administrative Divisions.	Codes and names for Canadian provinces, Mexican states, and states and state equivalents in the U.S.
	FIPS PUB 6-4, 1995 May	Counties and Equivalent Entities of the United States, its Possessions, and Associated Areas.	Codes and names for U.S. possessions and associated territories. USPS alphabetic postal codes for all U.S. entities.
Counties	FIPS PUB 6-4, 1995 May	Counties and Equivalent Entities of the United States, its Possessions, and Associated Areas.	Numeric codes and names for counties within states and state equivalents of the U.S., its possessions, and its associated territories.
EPA Region	U.S. EPA	Internal EPA Information.	Name of an EPA region (e.g., Region 9).
Tribal Land	Census Bureau, 1990	Department of Commerce, United States Bureau of the Census, Code 90.	Codes, names, and state-county locations of tribal lands.

Exhibit 1. Sources of SRT Data

7.0 CONCLUSIONS

OIRM has developed an initial set of SRTs for geopolitical units to be used by program systems whenever access to these reference data is required. The SRTs are accurate, current, and readily accessible. The SRTs benefit the Agency's goals for data standardization and integration. The tables are convenient and reliable for use by system developers, providing access to current and superseded data. Data integrity is maintained through ISN.

The initial set of SRTs is available to all EPA information system managers and system developers and can be accessed through mainframe and LAN connections. The content of the SRTs should be directly useful for many EPA system development and maintenance efforts. The initial utilization is voluntary and will be used to test and validate the design and access features implemented for the SRTs.

Future work on this project will develop notification and other procedures for change management. Future work will also extend the tables to a broader coverage to build a common core of key data elements used in EPA information systems.

The Points of Contact for obtaining more information about using the SRTs, whether for direct access or for implementation in another operational environment, are Larry Fitzwater or Jerry Widdowson, both at EIMD mail code (3405R). Both of them can also be contacted through EPA E-mail.